

Data Mining Project

Analysing AIr Quality

by Nishant Pal

# Abstract

# The project involves using weka tool to analyse air quality of five years from 2006-2010.Five data sets corresponding to each year have been used. The project tries to find the main reason that is contributing in changes of the air quality pattern in India. Classifying algorithms have been used in weka to predict the result with highest accuracy.

## Dataset description

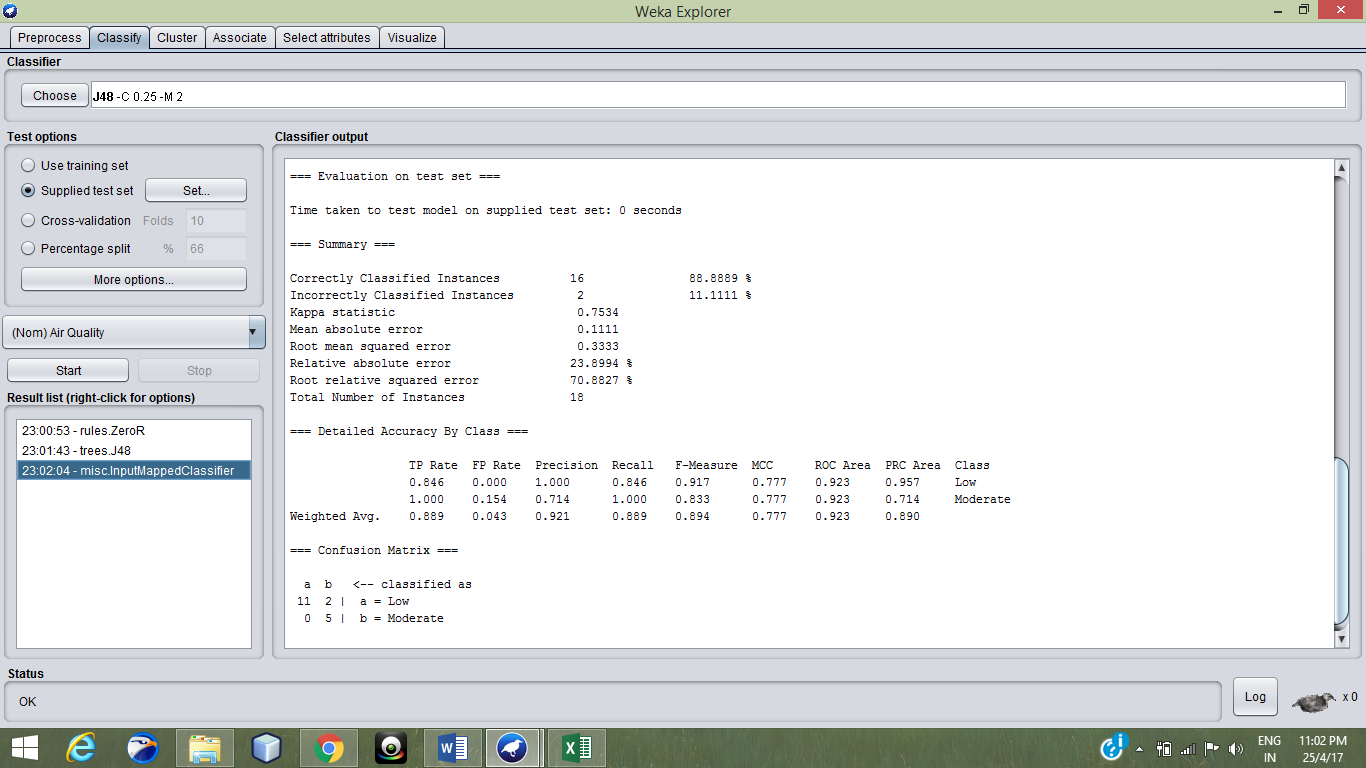
# Five air quality datasets from 2006-2010 have been analysed. The datasets are obtained from Open Government Datasets (OGD) platform in India. The datasets consist of basic attributes such as State, City, Location, Type of Area, Number of monitoring days(n),Ammonia(NH3)-Annual average(µg/m3) ,Standard Deviation, Percentage-exceedance(24 hourly),Air Quality. The ambient air quality monitoring network comprises of 342 monitoring stations covering 128 cities/towns of the country. The data is sourced from National ambient air quality status & trends in India

# Algorithms Used

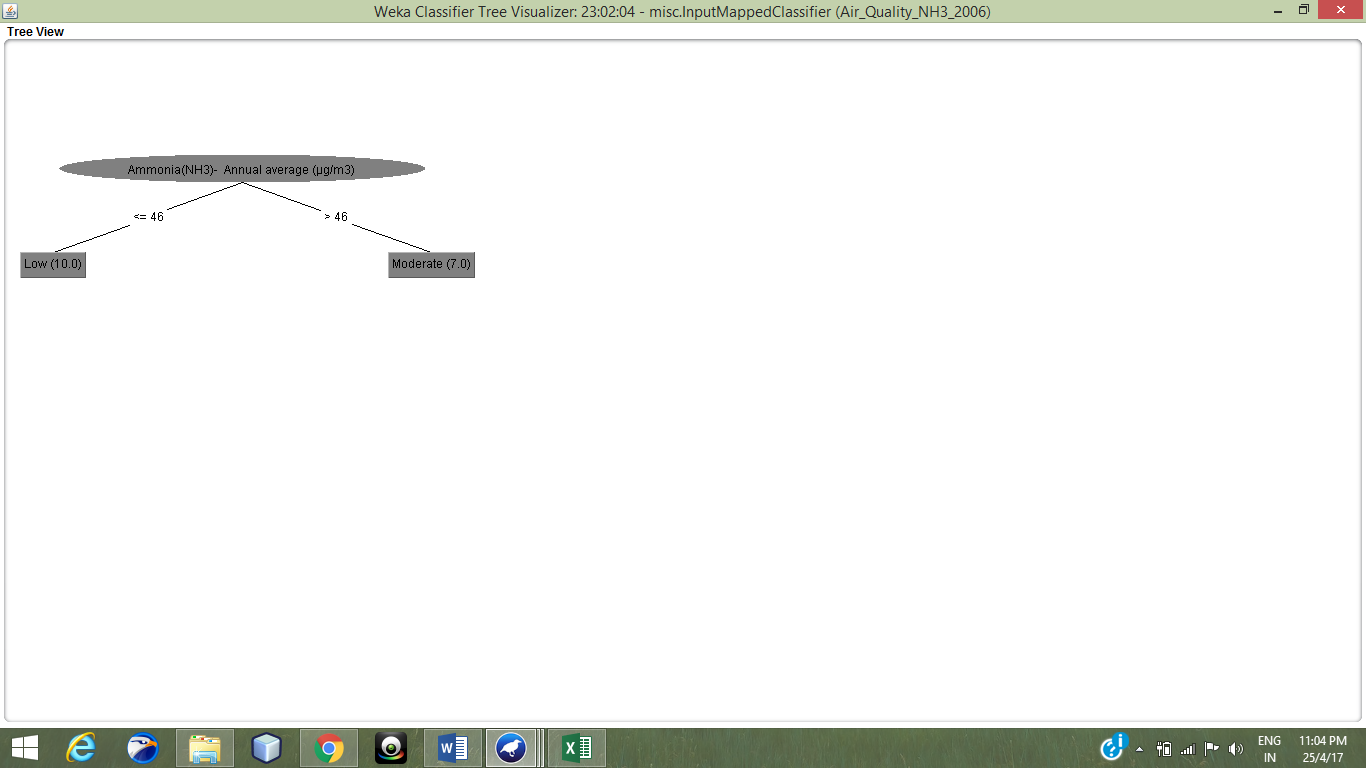
* J48
* LMT
* Decision Tree
* Decision Forest
* Decision Stump
* ReliefFAttribute Evaluator

Experiment

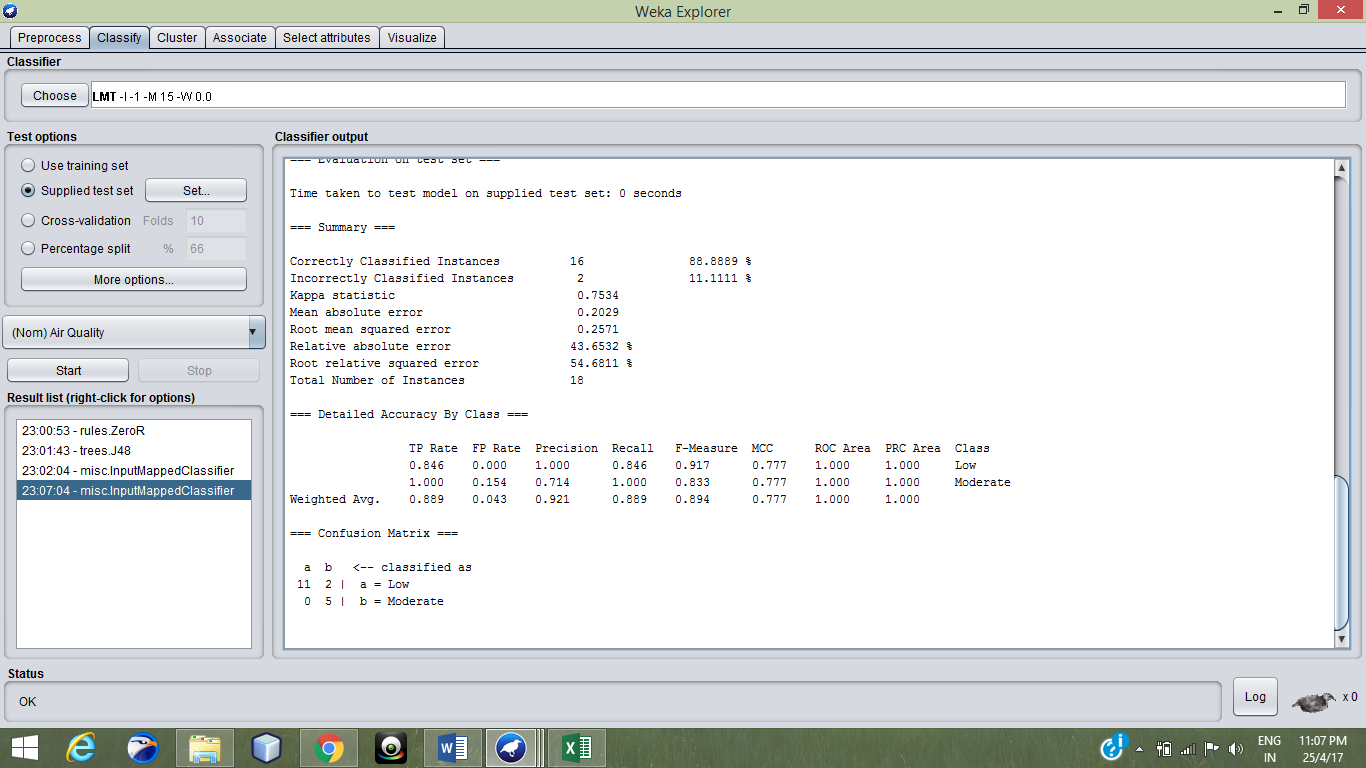
Using J48



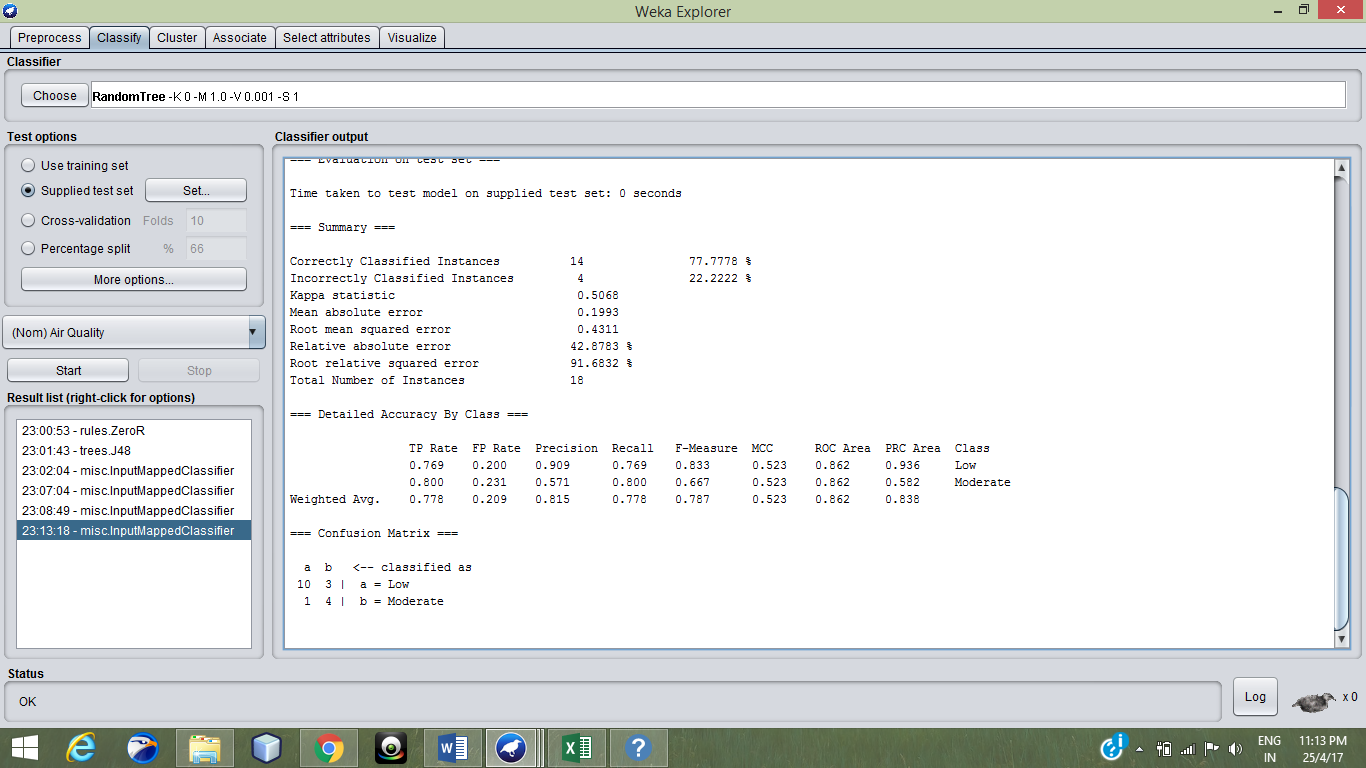
Tree Representation obtained



Using LMT



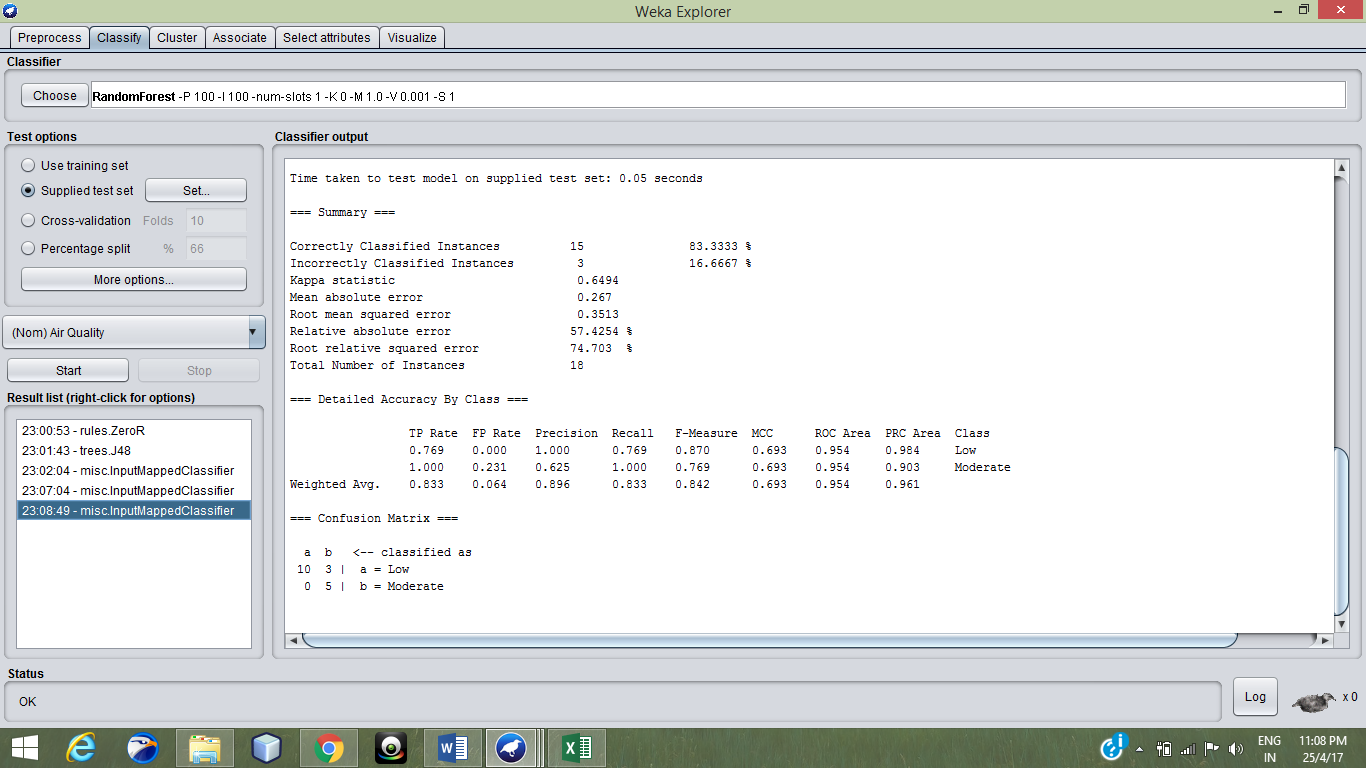
Using Decision Tree



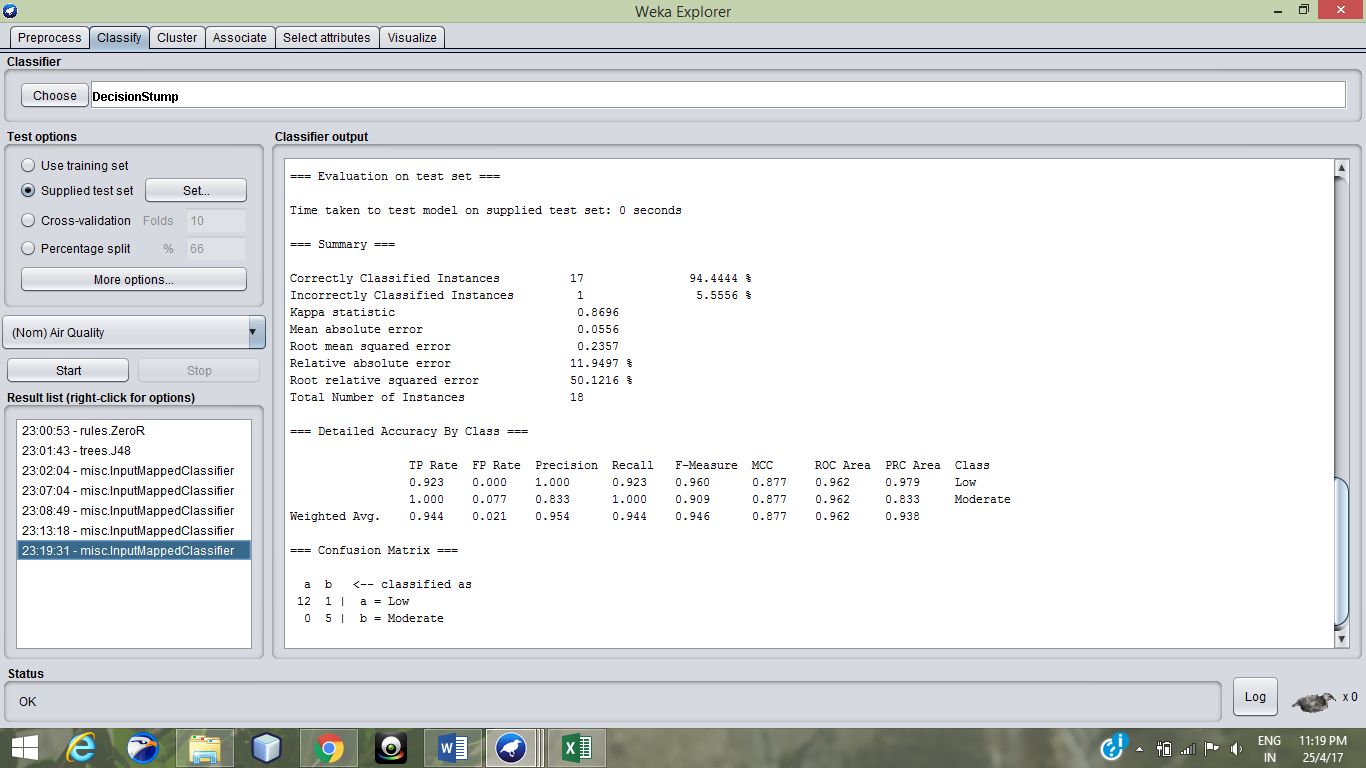
Tree Representation obtained



Using Random Forest



Using Decision Stump

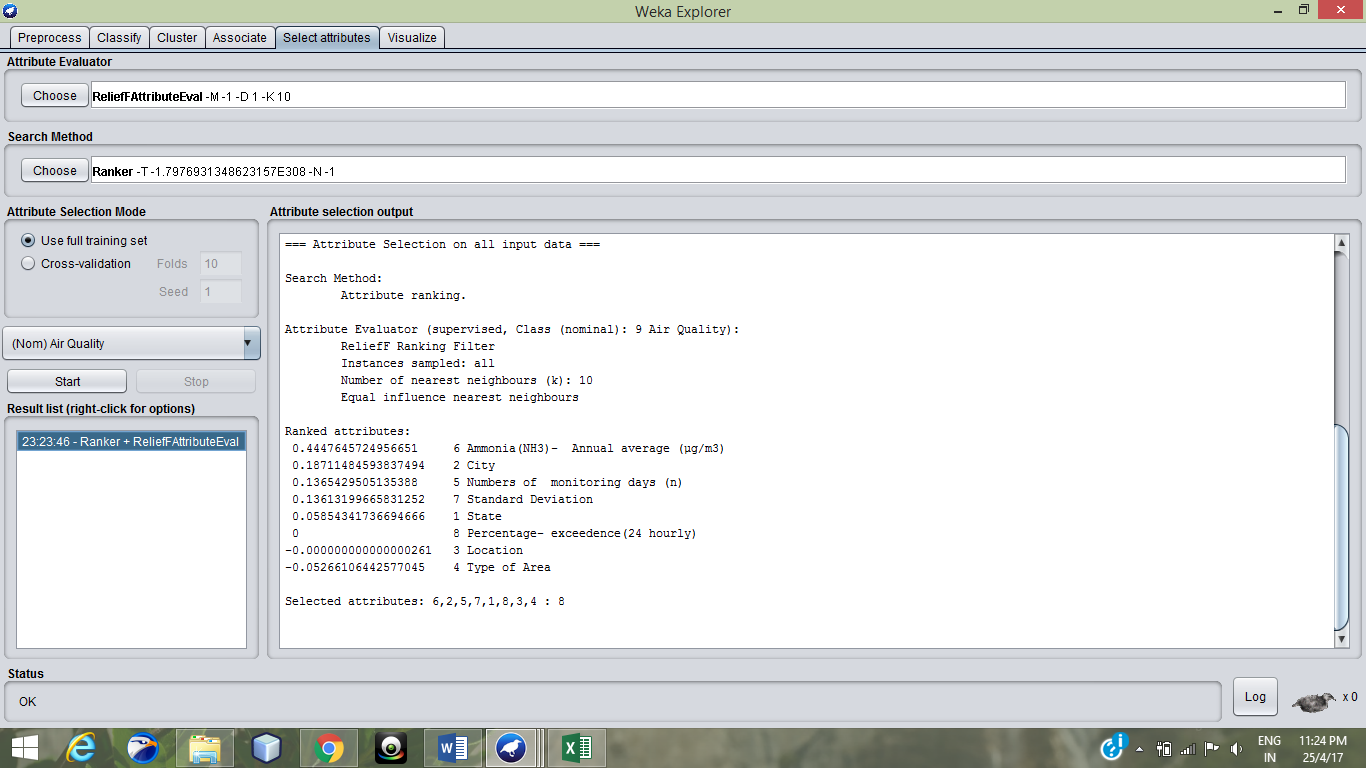


Highest accuracy obtained in this algorithm

Attribute Selection

Using ReliefFAttribute Evaluator

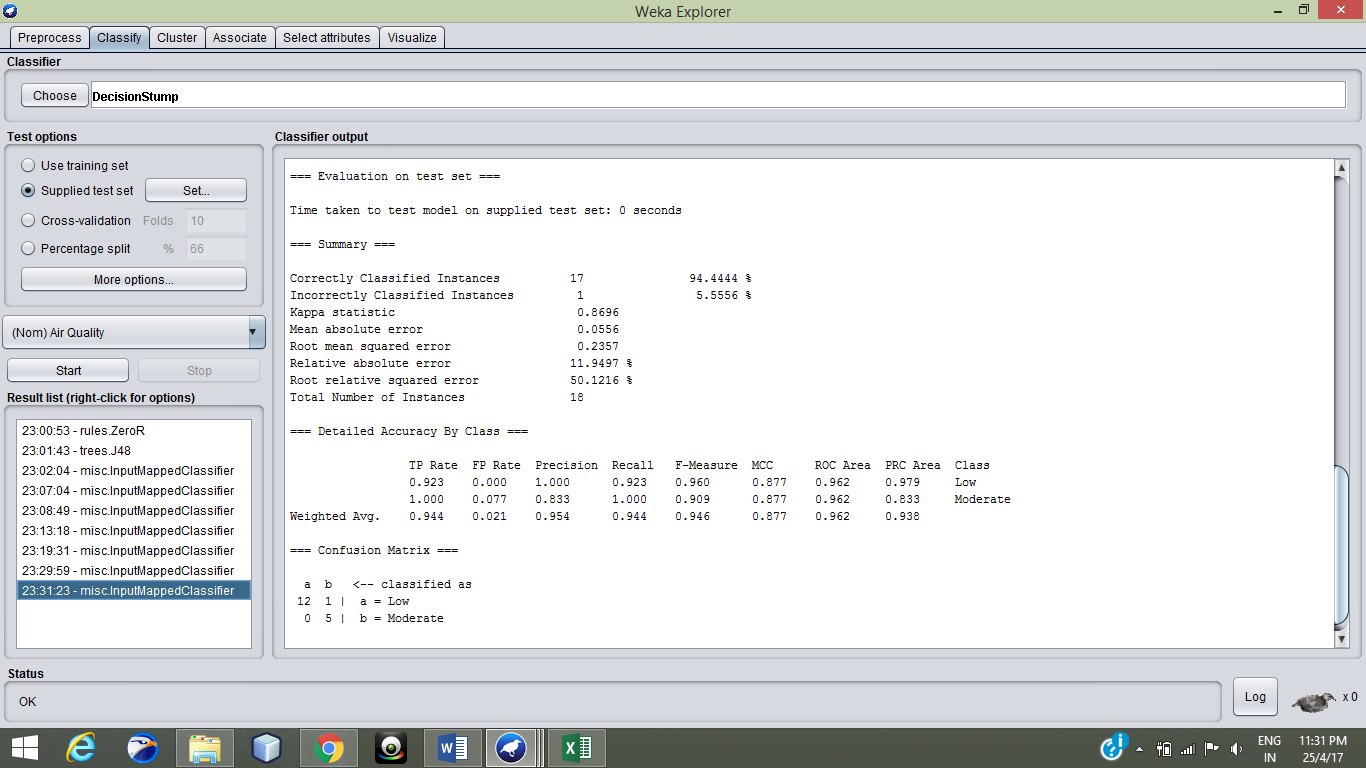
Using Ranker Method



Important attributes obtained:

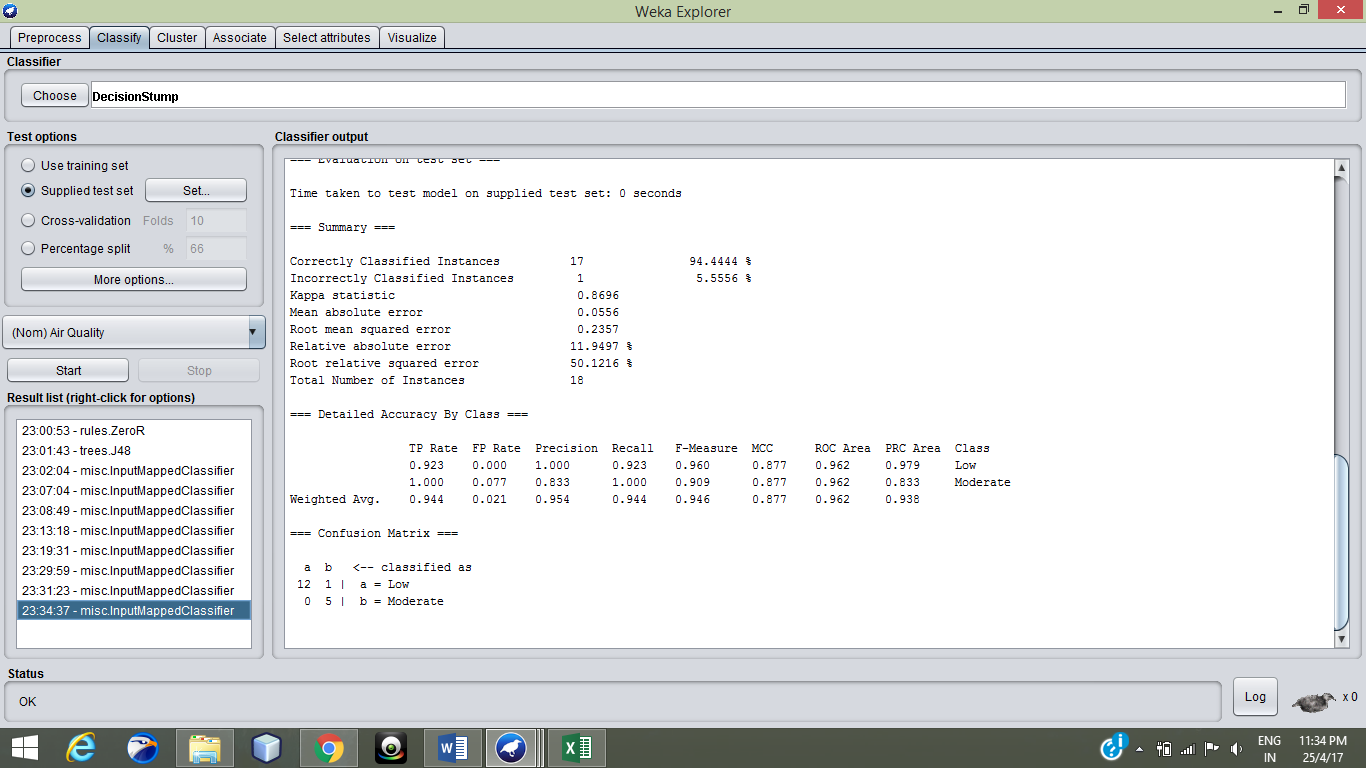
1. Ammonia (Nh3)-Annual average
2. City
3. Number of Monitoring days
4. Standard deviation
5. State

DecisionStump algorithm with selected attributes



Accuracy is still the same

DecisionStump algorithm with Ammoni(Nh3) -annual average attribute



Accuracy is still the same

Hence the major contributing attribute is :

Ammonia (Nh3)

Conclusion

Hence we conclude that the air quality is mostly affected by the ammonia concentration in that region. When the average ammonia concentration is less than 46µg/m3 ,the air quality is low. Whereas when the ammonia concentration is greater than

or equal to 46µg/m3 the air quality is moderate.